

## Notional Scoring for Technical Review Weighting As Applied to Simulation Credibility Assessment

*Joe Hale*

NASA

Marshall Space Flight Center

Huntsville, AL

256-544-2193

[joe.hale@nasa.gov](mailto:joe.hale@nasa.gov)

*Bobby Hartway*

AEgis Technologies Group

631 Discovery Drive

Huntsville, AL 35806

256-922-0802

[b.hartwar@aegistg.com](mailto:b.hartwar@aegistg.com)

*Danny Thomas*

AEgis Technologies Group

631 Discovery Drive

Huntsville, AL 35806

256-922-0802

[danny.thomas@aegistg.com](mailto:danny.thomas@aegistg.com)

Keywords:

“Modeling and Simulation Tools”, “NASA”, “Credibility Assessment”

**ABSTRACT:** *NASA's Modeling and Simulation Standard requires a credibility assessment for critical engineering data produced by models and simulations. Credibility assessment is thus a “qualifying factor” in reporting results from simulation-based analysis. The degree to which assessors should be independent of the simulation developers, users and decision makers is a recurring question. This paper provides alternative “weighting algorithms” for calculating the value-added for independence of the levels of technical review defined for the NASA Modeling and Simulation Standard.*

## INTRODUCTION

The primary goal of NASA-STD-(I)-7009, Standard for Models and Simulations, is to ensure that analysts properly report information that contributes to the credibility of results from models and simulations (M&S) to those making critical decisions [1]. The standard addresses development and application of M&S, as well as analysis, documentation, and presentation of the results from M&S. As determined by the Program for analysis that support “critical decisions,” it may apply to M&S used for Design and Analysis; Natural Phenomena Prediction; and Manufacturing, Assembly, Test, Operations and Evaluation. It may apply to all types,

sizes, and integration scales of M&S, from simple analytical spreadsheet models to extremely large, complex, distributed simulations for integrated systems simulation. It may apply to all scales of M&S application, from very quick-turnaround trade studies, to multiple program-phase use across years of program time.. The NASA Standard defines a one-dimensional, top-level scale for the uniform classification and reporting of M&S results credibility across all applications. The scale ranges from a perfect 4 down to 1. Evaluators have added a Level 0 to represent simulations that are too early in the development process to assess or simulations about which the evaluator has no information.

## LEVELS OF REVIEW

Analysts have long used peer reviews, independent assessments, expert opinions, user groups, panels, juries and the like to help establish the credibility of simulations. It is generally conceded that the quality of the review affects the credibility of the results. Decision makers have more confidence in a thorough independent review conducted by experts, for example. In the spirit of the Credibility Levels, this scale differentiates Levels of Review.

## EFFECTS OF REVIEW LEVEL ON CREDIBILITY LEVEL

Figure 1 presents the first algorithm for determining the effect of the level of review on the credibility of simulation results. This grid is very much like that used in RISK, wherein the two axes are LIKELYHOOD and CONSEQUENCE, and RISK is the interior grid.

		Review-Weighted Score				
Rigorous	4	2.0	2.5	3.5	3.7	4.0
	3	1.0	2.0	2.5	3.0	3.5
	2	0.5	1.0	2.0	2.2	2.5
	1	0.2	1.0	1.0	1.2	1.5
	0	0.0	0.3	0.5	.07	1.0
		0	1	2	3	4
		No Reviews	Informal Peer	Formal Peer	Formal Certified Peer	Formal Certified Independent

Scores nominal for the level of review  
 Scores improved by the level of review  
 Scores degraded by the level of review

**Figure 1: Heuristic Method**

For any given criterion evidence evaluation score, shown on the left hand side of the grid, each succeeding level of technical review either reduces or increases the resulting, weighted value of the criterion, as shown. The built-in rule is that a minimum level of tech review is required at each corresponding scoring level for assessed criterion evidence. A tech review at a level less than or equal to the evidence will reduce the weighted value score, and a tech review level greater than the evidence level will advance the weighted value score. The grid valuations are non-linear (off the diagonal). Reviews above the nominal level do not improve simulation credibility to the degree

- Level 4 – Formal external peer review accompanied by an independent evaluation of the evidence under review (e.g., independent reproduction of the relevant findings)
- Level 3 - Formal external peer review
- Level 2 – Formal internal peer review
- Level 1 – Informal internal peer reviews
- Level 0 – No reviews

that reviews below the nominal level decrease it. The perception of the evidence is obviously dependent on reviews. The problem is that the values are arbitrary.

The second method considered adds an absolute weighted score to the sub-factor. As shown in Figure 2. Technical Reviews again have values from 0 the 4. The analyst calculates the weighted score by multiplying a factor, 0.1 in this example, by the Technical review value and adding the result to the sub-factor. The problem with this approach is that a sub-factor could achieve a score greater than 4. Note that having no review of the evidence, or no evidence, produces a “not applicable” cell in this and the following method.

Subfactor					
4	N/A	4.1	4.2	4.3	4.4
3	N/A	3.1	3.2	3.3	3.4
2	N/A	2.1	2.2	2.3	2.4
1	N/A	1.1	1.2	1.3	1.4
0	0	N/A	N/A	N/A	N/A
Tech Rev	0	1	2	3	4

**Figure 2: Liner Compensation Method**

The last method is the currently selected method for the standard, which is undergoing final acceptance voting at the time of this writing. This method employs the same Analytical Hierarchy Process to roll up the technical review and the sub-factor as is used to roll up the sub-factors. The sub-factor and the technical review weights are normalized, i.e. they sum to one. This has the desired effect of augmenting or reducing the sub-factor score depending on the quality of the technical review. When sub-factor level and the Technical review level are equal



there is no effect. The weighting factor for technical review is constrained to no more than 30% of the weight or it would be possible for a technical review to raise or lower a sub-factor more than a whole level. The Responsible Party could apply a low, medium, or high

weight to the Technical Review, relative to the weight applied to the evidence, by using an Evidence/Tech Rev ratio of 90/10, 80/20, or 70/30, respectively. Figure 3 shows some examples of this approach.

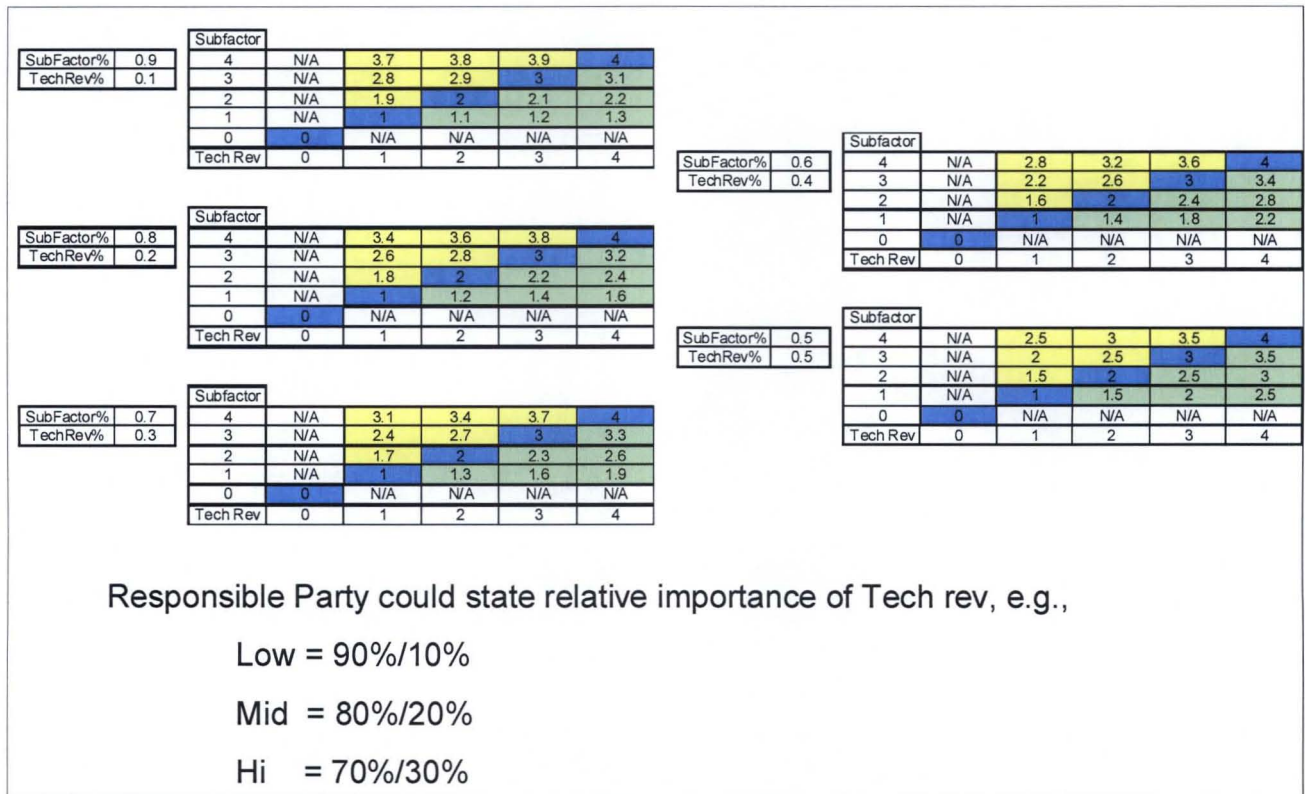


Figure 3: Examples of the Normally-Weighted, Constrained Method

## REFERENCES

- [1] NASA-STD (I) 7009, Standard for Models and Simulations.

This paper refers to the interim standard that NASA released to the public in December of 2006 and includes some of the thinking considered for the revision. The NASA process of reviewing interim standards requires a one year review. NASA has completed the revision and review process, scheduling the release of the final standard in April of 2008. At the time of this writing, NASA has not yet approved the final standard.

## AUTHOR BIOGRAPHIES

**JOSEPH PETER HALE, II** has worked with NASA at the Marshall Space Flight Center (MSFC) since 1985. He received a B.A. in Psychology from the University of Virginia in 1976 and an M.S. in Applied Behavior Science (Psychology) from Virginia Tech in 1981. He continued graduate studies at Virginia Tech through 1984 completing all coursework for a Ph.D. in Human Factors Engineering while participating as a Graduate Co-op Student at MSFC. In 1990, he received an M.S. in Systems Management from the Florida Institute of Technology.

Mr. Hale is currently Lead for the Ares Integrated Management of Modeling and Simulation. Prior to that, he was the Lead for the Exploration Systems Mission

Directorate's Integrated Modeling and Simulation Verification, Validation, and Accreditation activity. Much of his prior work at MSFC was as a Human Factors Engineer, working various projects, including Spacelab and the International Space Station. He spent five years as Team Lead for the Human Engineering and Analysis Team.

Mr. Hale is a Certified Human Factors Professional (CHFP) (Board of Certification in Professional Ergonomics), is a founding member and first president of the Tennessee Valley Chapter of the Human Factors and Ergonomics Society (HFES).

**BOBBY HARTWAY** is a Senior Research Scientist with AEGIS Technologies Group in Huntsville, Alabama. He has developed a new paradigm for simulation characterization and requirements development for space and defense systems. He is using this paradigm to support NASA's activities for integrated management of modeling and simulation. Mr. Hartway is a Certified Modeling and Simulation Professional.

**DANNY THOMAS** is a Senior Research Scientist with AEGIS Technologies Group in Huntsville, Alabama. He is supporting NASA's effort to institute consistent management practices for simulation development and use. He has developed simulations for space and defense. Mr. Thomas is a Certified Modeling and Simulation Professional.